Editorial
On line



SOCIETAT D'HISTÒRIA
NATURAL DE LES BALEARS

Eines genètiques i evolutives per a la descripció de la diversitat biològica

Joan PONS i Pere BOVER

Grup de Recerca Biodiversitat, Sistemàtica, i Evolució
Institut Mediterrani d'Estudis Avançats, IMEDEA (CSIC-UIB)
Miquel Marquès, 21
07190 Esporles (Illes Balears)
E-mail: jpons@imedea.uib-csic.es

L'increment de l'activitat antropogènica dels darrers dos segles ha provocat una dràstica reducció i fragmentació dels hàbitats naturals i la seva diversitat biològica, sobretot a dos dels punts calents de biodiversitat com són les conques Mediterrània i Amazònica (Kruess i Tscharntke, 1994; Myers *et al.*, 2000; Brooks *et al.*, 2006; Cardinale *et al.*, 2012). Aquest procés s'ha accelerat tant que la majoria de les espècies s'extingirà abans de que es puguin identificar i descriure ja que el nombre de taxònoms és baix i, a més, s'està disminuint, fenomen que s'ha denominat crisi taxonòmica (Tautz *et al.*, 2003). Aquest fenomen és particularment greu en els grups sistemàtics composats per milers d'espècies i on les descrites representen menys de la meitat de les estimades, com per exemple en els artròpodes (Odegaard, 2000). I el problema també es pot ampliar a nivell poblacional ja que la diversitat genètica és crucial per a la viabilitat de les espècies i per als processos d'especiació (Gugerli *et al.*, 2008).

Fa uns 30 anys, els estudis evolutius i sistemàtics a nivell molecular només es podien realitzar sobre organismes model com per exemple *Drosophila melanogaster*, *Caenorhabditis elegans* i *Mus musculus*, però dues tècniques revolucionàries varen permetre desenvolupar-los a nivell de qualsevol llinatge eucariota. Un dels mètodes fou l'amplificació *in vitro* de l'ADN mitjançant la Reacció en Cadena de la Polimerasa (sigles PCR en anglès, *Polymerase Chain Reaction*) que combina cadenes curtes d'ADN de cadena

simple (anomenats oligo-nucleòtids, iniciadors, encebadors o *primers*), una polimerasa termoestable i una repetició seqüencial de tres cicles de temperatura per multiplicar de forma exponencial el nombre de còpies d'un fragment d'ADN: la desnaturalització de la doble cadena d'ADN, la hibridació de l'encebador a l'ADN i polimerització de la cadena complementària (Mullis *et al.*, 1986). L'altre fou la possibilitat de desxifrar l'ordre dels nucleòtids d'una cadena d'ADN (seqüència nucleotídica) mitjançant l'aturada de la síntesi enzimàtica de la polimerasa degut a la incorporació de desoxinucleòtids marcats (Sanger *et al.*, 1977). Durant les dues darreres dècades, l'ús combinat d'ambdues tècniques ha permès obtenir les seqüències nucleotídiques d'espècies diferents per a reconstruir arbres filogenètics els quals mostren les relacions dicotòmiques ancestre-descendents d'aquestes fent possible contrastar-los amb la taxonomia clàssica basada en caràcters morfològics (Li, 1997). Fins fa relativament poc, el gens disponibles per a construir filogènies moleculars eren molt escassos perquè mancaven oligo-nucleòtids 'universals' per poder amplificar mitjançant PCR el mateix gen a vàries espècies. La majoria de seqüències disponibles estaven localitzades en el genoma mitocondrial (Simon *et al.*, 1994) i en el clúster de gens nuclears que codifiquen per l'ARN ribosomal (Hillis i Dixon, 1991).

Els mètodes de seqüenciació de nova generació (NGS, de l'anglès *Next Generation Sequencing*) desenvolupats ja dins el segle XXI per abaratir el projecte genoma humà han donat lloc a una segona revolució de la filogenètica molecular. Encara que avui en dia hi ha vàries plataformes de NGS (454, IonTorrent, Solexa, i PacificBio), el mètode de seqüenciació massiva més estès a l'actualitat és el de la plataforma Illumina bàsicament degut a l'alta qualitat i baix cost de les seves seqüències (Quail *et al.*, 2012). Per exemple, una carrera (o *run*) del seqüenciador Illumina model HiSeq 2500 produeix 330 milions de seqüències de 150 nucleòtids per uns 3.000 € i en tan sols una setmana. Aquest mètode es basa en un procés previ molt complex a micro-escala que comença tallant el genoma aleatòriament en fragments curts de 50-300 bases per després afegir-ne adaptadors en el seus extrems. Posteriorment, ja dins el seqüenciador, aquests fragments es fixen a la superfície d'un *microxip* de vidre recobert d'oligo-nucleòtids complementaris als adaptadors amb una disposició geomètrica. Una vegada fixats, s'amplifiquen enzimàticament mitjançant una estructura de pont per formar conjunts clonals (o *clusters*) de la seqüència primària i, així, augmentar la senyal de la reacció de seqüenciació. Finalment, la imatge presa a tot el *microxip* durant cada fase de síntesi permet la detecció de la incorporació, o no, d'un determinat nucleòtid a cada un dels milers de *clusters* del *xip* el qual s'enregistra a un ordinador per la seva anàlisi posterior. Pot consultar-se un video a (http://www.youtube.com/embed/HMyCqWhwB8E?iframe&rel=0&autoplay=1).

Gràcies al desenvolupament de NGS i d'altres tecnologies relacionades s'ha pogut passar, en poc més d'unes dècades, de construir filogènies d'un sol gen i pocs individus a arbres formats a partir de milers de seqüències genòmiques de moltes espècies (Kawahara i Breinholt, 2014; Misof *et al.*, 2014; Pecon-Slattery, 2014; Wickett *et al.*, 2014; Zhouq *et al.*, 2014).

El primer pas de qualsevol anàlisi filogenètica és construir un alineament múltiple de les seqüències obtingudes per a diferents espècies amb l'objectiu d'optimitzar globalment el nombre d'identitats nucleotídiques i així establir una hipòtesi primària de l'homologia de les posicions nucleotídiques examinades (Pons i Vogler, 2006). Aquest procés és molt ràpid malgrat el nombre de seqüències analitzades sigui considerable. Però si el conjunt de dades

és molt gran, sobre tot d'espècies, llavors reconstruir la topologia representa un repte computacional. Els primers mètodes filogenètics es basaven en mètodes de distàncies genètiques, però avui en dia ja no s'empren perquè presenten vàries limitacions metodològiques (Hillis *et al.*, 1996). Les anàlisis basades en el criteri de parsimònia foren el principal mètode de reconstrucció d'arbres filogenètics fins fa una dècada perquè el seu càlcul computacional era molt ràpid. Aquest mètode consta de múltiples iteracions de dues etapes. A la primera, a partir d'un primer arbre generat hi ha una poda d'un petit conjunt de branques (subarbre) que es recol·loca a un altre punt de l'arbre, i una segona, on es compten el nombre de canvis nucleotídics informatius (caràcters derivats compartits) a la topologia de l'arbre (Hillis *et al.*, 1996; Li, 1997). A cada iteració, la nova topologia reordenada és només acceptada i guardada en memòria si té un nombre menor de canvis o igual que l'anterior. Generalment, s'empra una aproximació heurística amb un nombre fixat d'iteracions que comencen des d'una topologia construïda a l'atzar per a evitar quedar atrapat en un solució subòptima de l'espai d'arbres ja que és impossible comprovar els bilions de topologies possibles (Hillis *et al.*, 1996). Recentment, l'augment de poder computacional ha permès implementar mètodes més complexos basats en criteris de màxima versemblança i baiesians. Aquests mètodes, a l'hora de realitzar l'anàlisi filogenètica, optimitzen cada tipus de substitució de forma independent així com la velocitat de canvi de cada posició nucleotídica o aminoacídica (Felsenstein, 2004). Breument, la màxima versemblança cerca aquells valors per als paràmetres del model evolutiu seleccionat que maximitzen la probabilitat de les seqüències observades, mentre que el mètode baiesià tracta els paràmetres com variables a l'atzar amb una distribució coneguda que s'afinen amb les dades observades (Felsenstein, 2004). A més, aquesta aproximació baiesiana permet l'avaluació de l'interval de confiança de cada un dels paràmetres del model evolutiu implementat, així com la comparació explícita de models amb un gran nombre de paràmetres.

Els arbres filogenètics, a més de resoldre les relacions evolutives entre tàxons, pot incloure un marc temporal que permeti valorar la congruència dels diferents canvis evolutius amb determinats esdeveniments geològics. Als anys 60, el treball de Zuckerkandl i Pauling (1962) introduí el terme rellotge molecular que suggeria que les mutacions del gens sota selecció neutra, per exemple els marcadors filogenètics, eren relativament constants al llarg del temps. No obstant això, estudis posteriors demostraren que aquesta constància era poc comuna i, per tant, es desenvoluparen mètodes de relaxació de rellotge molecular que permeten que la taxa de substitució nucleotídica varií a les diferents branques de l'arbre (Thorne *et al.*, 1998; Sanderson, 2002; Drummond *et al.*, 2006). El desenvolupament d'aquestes tècniques moleculars i aquests mètodes filogenètics entre d'altres, ha tengut un gran impacte en l'estudi de la biodiversitat tant a nivells taxonòmics com sistemàtics de gran llinatges evolutius o a nivell d'espècie, inclús d'aquelles extintes. Alguns d'aquests temes seran desenvolupats en les properes seccions.

Estudis de biodiversitat basats en el codi de barres de l'ADN

A principis d'aquest segle XXI, el Professor Hebert proposà l'ús de la seqüència d'ADN de la regió 5' del gen mitocondrial citocrom oxidasa I (*cox1*) com identificador de cada

espècie perquè aquest conté mutacions diagnòstiques (Hebert *et al.*, 2003). Es va observar que aquesta especificitat entre morfologia i seqüència *cox1* a nivell d'espècie era similars a la que es podia trobar entre un producte d'un supermercat i el seu codi de barres, i per això, aquest mètode s'anomenà codi de barres d'ADN (*DNA barcoding*). Els avantatges d'aquesta nova aproximació a la identificació d'espècies, conjuntament amb la conscienciació de les problemàtiques de l'extinció accelerada de la biodiversitat i la crisi taxonòmica, provocaren que es creàs el Consorci del Codi de Barres de la Vida (CBOL, de l'anglès *Consortium for the Barcode of Life*) per promoure entre la comunitat científica l'ús del codi de barres d'ADN com a sistema de reconeixement i identificació ràpida de les espècies. Aquest consorci ha establert normes i protocols per poder assignar correctament la seqüència d'un nou espècimen a una espècie o identificar-la com a pertanyent a una espècie desconeguda comparant les agrupacions basades en taxonomia clàssica amb els clústers de seqüències de *cox1* definits mitjançant algoritmes filogenètics (Hebert *et al.*, 2003). El gen elegit a animals fou el *cox1* però com aquest era ineficient a plantes es varen seleccionar els gens cloroplàstics subunitat llarga de la RuBisCo (*rbcL)* i maturase K (*matK),* i a fongs el gen nuclear espaiador transcriptor intern de l'ARN ribosomal (*ITS)*. Estudis posteriors corroboraren els diversos avantatges dels codis de barres d'ADN: la universalitat i l'homologia de l'ADN, en contraposició a l'ambigüitat i exclusivitat la de la morfologia. La seva precisió taxonòmica incorpora informació sistemàtica, automatització i baix cost que faciliten la seva aplicació en contextos diversos (fins i tot per als no especialistes). A més, es caracteritza per la portabilitat, l'accés rutinari i immediat a la informació, i la utilitat en un ampli espectre filogenètic i taxonòmic dels organismes (Hebert *et al.*, 2003; Hajibabaei *et al.*, 2005). No obstant això, aquest mètode també mostra punts febles: establir el valor òptim de tall a nivell de divergències genètiques per delimitar espècies és problemàtic ja que la taxa de substitució nucleotídica és molt variable dins i entre llinatges, fins inclús ser insuficient per discriminar espècies estretament relacionades. També cal esmentar que els grups genètics basats en la història evolutiva d'un sol gen no poden inequívocament traduir-se en la història de les espècies, i finalment i més important, el coneixement taxonòmic construït en els últims tres segles a partir de Linneo no es pot incorporar al nou marc (Pons *et al.*, 2006; Rubinoff *et al.*, 2006).

L'èxit inicial dels estudis basats en codis de barres a nivell local va fer que diverses institucions internacionals posassin en marxa projectes a gran escala que incloguessin la seqüenciació de milers de codis de barres de l'ADN per a estudiar grups taxonòmics complets o extenses regions geogràfiques: FISH-BOL (la totalitat d'espècies de peixos a escala global), ABB Iniciativa (aus), Plant DNA Barcode Project (plantes), Fungal Barcode (fongs), INBIPS (espècies invasives i paràsits), i BioCode Moorea (totes les espècies macroscòpiques de Moorea). El gener de 2015 al web http://www.barcodinglife.org/ de la Universitat de Guelph hi havia depositats i a disposició pública 3.676.066 codis de barres que provenen en la seva majoria d'espècies col·lectades en zones desenvolupades com Europa, Nord-Amèrica i Austràlia. Per exemple, hi ha depositats els codis de barra de 29.992 aus de 192 països [4.778 espècies putatives (descrites + críptiques + desconegudes) de les que 4.080 estan identificades taxonòmicament], 61.846 mamífers de 164 països (3.224 / 2.108), 2.380.002 d'insectes de 216 països (80.392 / 260.676), 79.829 mol·luscs de 156 països (12.658 / 8.840), i 4.740 platihelmints de 85 països (568 / 670). Si analitzam aquestes dades podem concloure que a aus i a mamífers el nombre d'espècies putatives és

molt similar a les descrites, però en canvi el nombre de desconegudes i críptiques a mol·luscs i, sobretot a insectes, superen en gran quantitat a les descrites. Finalment, els resultats a platihelmints suggereixen un gran nombre d'errors a l'hora d'identificar les espècies morfològicament.

El codi de barres d'ADN també té altres aplicacions molt interessants com són l'anàlisi d'aliments processats o de restaurants per saber de quines espècies animals o vegetals estant fets i així evitar el frau (Lowenstein *et al.*, 2009; Di Pinto *et al.*, 2013; Vartak *et al.*, 2015), saber de quin tipus de plantes està feta la mel (Bruni *et al.*, 2015), inferir la dieta d'una espècie a partir del contingut estomacal o mostres fecals (Pons, 2006; Jurado-Rivera *et al.*, 2009; Ibanez *et al.*, 2013), identificar comerç il·legal d'espècies CITES (Liu *et al.*, 2013), detecció d'ous i larves de paràsits (Gariepy *et al.*, 2014), o identificar a quines espècies piquen els hematòfags (Pettersson *et al.*, 2013). Finalment, l'aplicació dels avenços generats per les plataformes NGS al DNA barcoding permetran que es puguin estudiar comunitats completes en lloc d'individu per individu com es fa actualment i, particularment, a aquells de mida petita (McMahon *et al.*, 2014).

Filogenòmica de Metazous

Un dels enigmes que més ha captivat a zoòlegs i biòlegs evolutius és quan i perquè apareix la multi-cel·lularitat i si aquest tret només ha esdevingut una vegada durant l'evolució o bé ho ha fet vàries vegades de forma independent. Dos estudis recents han intentat resoldre aquesta pregunta mitjançant la transcriptòmica, és a dir, la seqüenciació de milers de gens que són expressats a un teixit o un individu (Dunn *et al.*, 2008; Ruiz-Trillo *et al.*, 2008). En aquests procés, els ARN missatgers es retro-transcriuen *in vitro* a ADN complementari, s'amplifiquen per PCR i finalment són seqüenciats pel mètode de Sanger, i més recentment per plataformes NGS. L'objectiu de l'estudi del Dr. Ruiz-Trillo era conèixer quin és el grup (o grups) eucariota unicel·lular germà dels metazoous. Per analitzar-ho va seqüenciar 110 gens nuclears (63.000 pb, parells de bases) i 13 mitocondrials (7.800 pb) de 12 espècies de metazous de diversos grans grups com porífers, anèl·lids, insectes i vertebrats. Per les anàlisis filogenètiques s'empraren mètodes de màxima versemblança i baiesians a nivell de proteïna per així evitar la saturació deguda a les substitucions nucleotídiques múltiples que va fer que el conjunt de dades es reduís a 21.000 aminoàcids pels gens nuclears i 2.600 pels mitocondrials (Fig. 1). A més, l'estudi va incloure quatre grups de protistes dels quals s'ha suggerit que poden estar estretament emparentats amb Metazoa (el coanoflagel·lat *Monosiga ovata*, els ictiosporis *Sphaeroforma arctica* i *Amoebidium parasiticum*, i l'ameba *Capsaspora owczarzaki*), i finalment dues espècies del grup Amoebozoa (*Acanthamoeba castellanii* i *Mastigamoeba balamuthi*) que és l'evolutivament més llunyà de tots els altres esmentats anteriorment. Els 'arbres filogenètics nuclear i mitocondrial resultaren congruents i varen indicar que el grup més proper als metazous són els coanoflagel·lats seguits després de *Capsaspora* i Ichthyosporea. Aquests resultats no varen resultar del tot sorprenents ja que aquests grups primitius unicel·lulars com *Capsaspora* tenen al seu genoma gens homòlegs al gen *membrane-associated guanylate kinase* (*MAGI*) els quals codifiquen per proteïnes que tenen una funció de comunicació entre cèl·lules i per tant amb la multi-cel·lularitat.

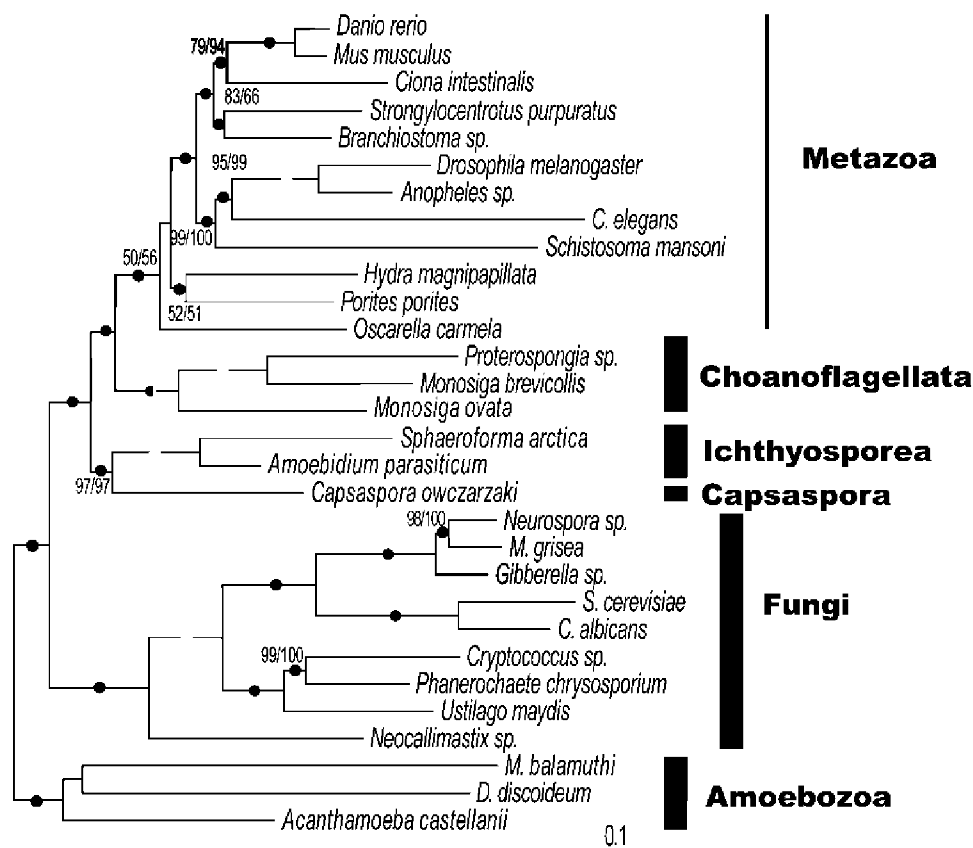


Fig. 1. Arbre filogenètic dels Opisthokonta basat en l'anàlisi baiesiana de 110 seqüències de proteïnes nuclears. Els punts negres en els nodes indiquen els valors de probabilitat posterior de 1.0, i els grisos si els valors de suport de *bootstrap* de màxima versemblança són de també del 100%. Modificat de Ruiz-Trillo *et al.* (2008).

Fig. 1. *Phylogenetic tree of the Opisthokonta based the Bayesian analysis 110 nuclear proteins sequences. Black dots on nodes indicate posterior probability values of 1.0, and grey ones if maximum likelihood bootstrap analysis also yields 100% support. Modified from Ruiz-Trillo* et al.*, 2008.*

Per altra banda, el principal objectiu del segon estudi esmentat (Dunn *et al.*, 2008) va ser resoldre les relacions filogenètiques entre els principals llinatges taxonòmics i els plans corporals dels metazous (Fig. 2). El conjunt de dades d'aquest estudi va estar compost per 150 gens de 71 espècies de metazous juntament amb sis grups externs que s'aconseguiren del transcriptoma de cada una d'elles. El projecte generà en total 40.000.000 pb (40 Mb) de dades d'ADN. Els arbres filogenètics resultants corroboraren la posició de molts dels grups taxonòmics descrits en base a estudis morfològics, és a dir, varen ser monofilètics i amb alt suport, com per exemple Metazoa, Deuterostomia (Chordata, Hemichordata + Echinodermata) i Protostomia. Però, per una altra banda, dins aquest darrer grup sorgiren

agrupaments diferents als descrits prèviament per la sistemàtica clàssica, com per exemple el nou grup anomenat Ecdysozoa que es caracteritza pel tret comú de tenir el cos recobert per una cutícula. Un altre exemple és Lophotrochozoa, grup que està format per espècies que tenen lofòfor (conjunt de cilis en forma de ventall al voltant la boca) típic de Phoronida i Brachiopoda o presenten una forma larvària anomenada trocòfora (larva amb dues bandes de cilis enmig del cos) característica de Mollusca, Anelida, Echiura, Sipuncula i Nemertea. Aquests resultats també mostraren que els celomats i mandibulats no són monofilètics i, per tant, la seva sistemàtica i taxonomia s'havia de reescriure. Una altra troballa interessant va ser que el grup basal a Metazoa era Ctenofora i no Porifera i Cnidaria tal com era d'esperar, i que els coanoflagelats era el grup germà dels metazous corroborant així l'estudi del Dr. Ruiz-Trillo. Aquesta aproximació filogenòmica basada en la transcriptòmica també s'ha aplicat amb èxit per resoldre les relacions evolutives entre els insectes (Misof *et al.*, 2014), plantes terrestres (Wickett *et al.*, 2014), mamífers (Zhouq *et al.*, 2014), papallones i arnes (Kawahara i Breinholt, 2014), i primats (Pecon-Slattery, 2014).

Aproximació filogenòmica a la diversificació global de les aus

Les aus han captivat des de sempre l'interès dels naturalistes perquè presenten trets inusuals i atractius com la capacitat de vol, la bellesa i coloració del seu plomatge, el seu cant harmoniós, la capacitat de llenguatge i el complex comportament de festeig, social i reproductor. Aquest grup d'animals s'ha estudiat àmpliament per part dels biòlegs per a intentar comprendre l'evolució, entre d'altres, de les característiques abans esmentades, però també per conèixer el seu patró de diversificació a través del temps i contrastar-lo amb el seu registre paleontològic (Brown *et al.*, 2008; Pacheco *et al.*, 2011). Per exemple, no està clar si la radiació de les Neoaves (el que serien els ocells moderns excepte els Paleognathae i Galloanseres) va ser gradual o explosiva (Pacheco *et al.*, 2011). Estudis previs havien fracassat en donar respostes clares a algunes d'aquestes preguntes bàsicament perquè les relacions filogenètiques no estaven completament resoltes i suportades degut a que les seqüències d'ADN obtingudes no eren suficientment informatives. Recentment, Jarvis *et al.* (2014) han abordat aquest problema mitjançant la seqüenciació dels genomes complets de 48 espècies d'aus que representen els principals llinatges taxonòmics. El primer pas fou seleccionar les seqüències homòlogues en cada genoma per a construir el conjunt final de dades que incloïa el 40% del proteoma (gens que codifiquen per proteïnes però que només representen un 3,5% de les seqüències d'una au), regions intròniques no codificants de 2.516 gens, i 3.756 elements d'ADN no codificant ultraconservats incloent 1.000 pb de seqüències flanquejants. El conjunt de dades total fou d'uns 42 milions de parells de bases d'ADN (42 Mb). La datació dels nodes de la filogènia d'aus va requerir la transformació del nombre de substitucions de nucleòtids a una estimació de milions d'anys. Per això, es restringí l'edat de 19 nodes de l'arbre filogenètic a l'edat de fòssils d'aus concrets com a valor mínim, i a més es va fixar l'edat per a alguns grups externs i una edat mínima per als ocells moderns (Neornithes) de 99,6 Milions d'anys [Ma]. Per a calcular aquesta anàlisi filogenètica s'empraren nou super-ordinadors durant unes setmanes ja que si s'hagués emprat un ordinador convencional, el temps de computació per acabar la tasca hagués estat de 400 anys. Els resultats suggereixen que Palaeognathae i Neognathae, que

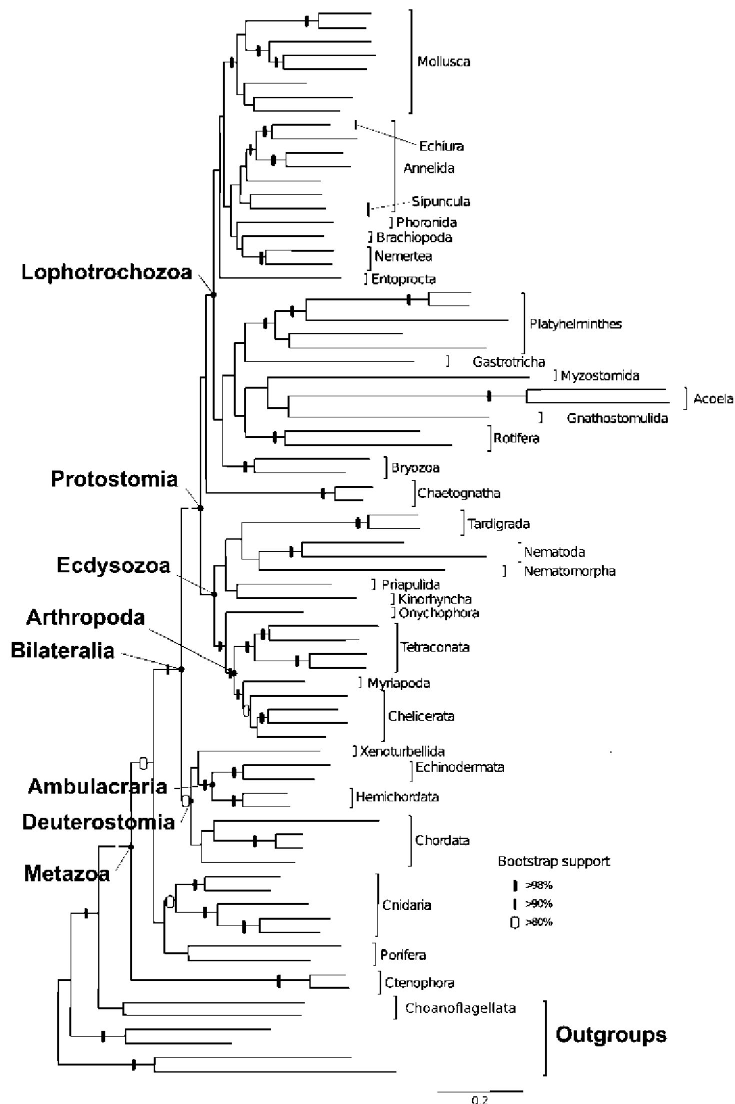


Fig. 2. Arbre filogenètic de 71 espècies de metazous, incloent membres dels llinatges taxonòmics més importants, a més de 6 grups externs basats en l'anàlisi de màxima versemblança de 42 Mb de seqüències de proteïnes. Branques amb diferents símbols indiquen els valors de suport de *bootstrap* que van des de 80 a 100%. Modificat de Dunn *et al.* (2008).

***Fig. 2**. Phylogenetic tree of 71 metazoan species, including most important taxonomic lineages, plus 6 outgroups based on maximum likelihood analysis of 42 Mb of protein sequences. Branches with different symbols indicates bootstrap support values ranging from 80 to 100%. Modified from Dunn* et al. *(2008).*

són els dos grans llinatges d'aus, es separaren en el Cretàcic Tardà (fa uns 100 Ma) i que l'edat estimada de la divisió basal dels Passeriformes, els quals representen un 60% de les 10.400 espècies d'ocells, és del voltant de 39 Ma. Aquesta filogènia robusta ha permès estudiar, entre d'altres, l'aparició del tret del llenguatge a les aus i que aquest s'ha desenvolupat independentment diverses vegades durant la història evolutiva d'aquests vertebrats. També s'ha pogut demostrar que les aus de presa són polifilètiques ja que aquestes són basals als grups germans Australaves i Afroaves. Aquest estudi també ha demostrat que grups taxonòmics definits per un caràcter o comportament comuns en realitat no són monofilètics i que aquestes característiques haurien evolucionat independentment vàries vegades i serien, per tant, casos de convergència evolutiva. Com exemples, es pot esmentar el tret de busseig per propulsió amb els peus que presenten espècies de cabussons (Podicipediformes, Columbea) i corbs marins (Suliformes, Passerea), o el tret de l'alimentació mentre es camina per zones d'aigües poc profundes que es pot observar a espècies de flamencs (Columbea) i a ibis i agrons (Passerea). Finalment, un dels resultats més interessants d'aquest estudi fou que la radiació de la majoria de les espècies actuals (Neoaves) fou explosiva i en només 10-15 milions d'anys coincidint amb la transició Cretàcic-Paleogen (K-Pg). Aquest moment coincideix amb l'extinció del seu grup germà, els dinosaures, que provocà que molts de nínxols ecològics quedassin disponibles.

Espècies extintes i l'ADN antic

Una altra de les disciplines de la genètica molecular que ha sofert un període d'expansió en els darrers 30 anys és, sense cap dubte, el de l'ADN antic, és a dir, l'anàlisi de seqüències d'ADN provinents de mostres d'espècies fòssils o sub-fòssils. El desenvolupament d'aquesta disciplina ha estat extraordinàriament ràpid, ja no tan sols pel que fa a les tècniques i mètodes de laboratori sinó també pel que fa als problemes que es volien abordar. Si inicialment la pregunta que es volia respondre era bàsicament si es podia obtenir ADN d'espècies extintes i veure quines eren les seves possibles relacions filogenètiques, la complexitat de les incògnites que es poden resoldre actualment s'ha incrementat de forma considerable: estudis poblacionals al llarg del temps, obtenció de genomes complets, etc.

Des de la recuperació als principis dels 80 d'ADN antic de quagga, un membre extingit de la família dels cavalls (Higuchi *et al.*, 1984), i l'obtenció d'ADN de restes humanes momificades (Pääbo, 1985), han estat molts els investigadors que han realitzat aportacions rellevants al camp de la genètica molecular fòssil. Després d'aquestes primeres troballes, la disciplina de l'ADN antic es va beneficiar del desenvolupament de les tècniques de PCR. Inicialment aquests treballs es centraren en l'obtenció de fragments d'ADN mitocondrial per a realitzar estudis filogenètics degut, bàsicament, a que l'ADN es degrada en fragments petits al llarg del temps i el nombre de mitocòndries per cèl·lula eucariota és gran (hi ha entre 100 i 10.000 còpies de genomes mitocondrials, el qual es transmet per via materna sense recombinació). Aquest va ser el cas, entre d'altres, dels moas de Nova Zelanda (Cooper *et al.*, 1992) i els llops marsupials de Tasmània (Thomas *et al.*, 1989). S'ha d'esmentar aquí que durant aquests primers estudis ja es va poder constatar la importància potencial que podria tenir aquesta recerca en camps com la paleontologia, arqueologia,

antropologia o inclús en la conservació d'espècies. No cal dir que els inicis de la disciplina no varen ser immunes als problemes derivats d'una de les característiques que més dificulten i sense cap dubte marquen la validesa dels resultats en ADN antic, la contaminació per part d'ADN exogen. Així, treballs que asseguraven que es podia obtenir ADN de mostres extraordinàriament antigues, com provinents de dinosaures (Woodward *et al.*, 1994; An *et al.*, 1995; Li *et al.*, 1995), de restes de plantes miocèniques (Golenberg *et al.*, 1990; Soltis *et al.*, 1992) o dels famosos insectes englobats dins ambre (Cano *et al.*, 1993), han estat posteriorment posats en dubte i rebutjats (veure, per exemple, Hedges i Schweitzer, 1995; Zischler *et al.*, 1995; Ward *et al.*, 1997; Penney *et al.*, 2013). En aquest sentit s'ha anat avançant en l'elaboració de protocols cada vegada més estrictes pel que fa a mecanismes per evitar contaminacions i s'establiren pautes de control en els mètodes d'extracció tal com la repetició de treballs en dos o més laboratoris diferents per a garantir la veracitat i fiabilitat dels resultats obtinguts, etc (veure per exemple, Cooper i Poinar, 2000). A mesura que s'han anat realitzant estudis a partir de mostres antigues s'han anat trobant dificultats així com les seves possibles solucions. Cal destacar com exemples d'aquests problemes la fragmentació de l'ADN antic, i els nombrosos canvis de C a T i de G a A que s'observen en seqüències antigues quan es comparen amb seqüències actuals (fenomen produït per desaminació de la citosina). Curiosament aquestes característiques resulten actualment importants per a verificar que les seqüències obtingudes són efectivament antigues (e.g., Briggs *et al.*, 2007).

Així, tots els avanços, entre d'altres, en mètodes d'extracció, clonació, elaboració de biblioteques genètiques i enriquiments de mostres per hibridació amb sondes d'ARN ha permès, juntament amb el desenvolupament de les tècniques de seqüenciació massives (les abans esmentades plataformes NGS), obtenir primerament genomes mitocondrials complets de diverses espècies animals. Aquest és el cas de, per exemple, moas de Nova Zelanda (Cooper *et al.*, 2001; Haddrath i Baker, 2001), *Mammuthus primigenius* (Krause *et al.*, 2006; Rogaev *et al.*, 2006), *Ursus spelaeus* (Bon *et al.*, 2008; Krause *et al.*, 2008), *Coelodonta antiquitatis* (Willerslev *et al.*, 2009) o *Bos primigenius* (Edwards *et al.*, 2010).

Però si fa uns 10 anys resultava impensable la seqüenciació de genomes complets d'espècies o grups taxonòmics extingits, especialment pels que fa a la línia evolutiva humana, a l'actualitat s'han aconseguit una sèrie de resultats que fan pensar en un futur més que prometedor. Fa uns anys, el cost tant en temps com en recursos econòmics per a l'obtenció de genomes complets o gran part de genomes era extraordinari. A més, si tenim en compte que hi ha molt poques còpies de gens nuclears per cèl·lula, es necessitava una quantitat important de material que moltes vegades no està disponible en determinades mostres fòssils. Després d'extreure l'ADN de la mostra, s'havia d'amplificar mitjançant una sèrie de PCR i els amplicons obtinguts s'havien de clonar per via de vectors bacterians, i els centenars o milers de clons s'havien de seqüenciar individualment per via del mètode Sanger. Ja per l'any 2005, l'aparició de les PCR d'emulsió de forma que cada molècula d'ADN és amplificada dins d'una gota d'emulsió oli-aigua i la seqüenciació amb la plataforma 454 Life Sciences (posteriorment adquirida per la companyia Roche) la qual permetia centenars de mils de reaccions de seqüenciació en paral·lel va suposar una alternativa molt efectiva a la clonació. El posterior desenvolupament de plataformes de seqüenciació massiva com Solexa (més tard adquirida per Illumina), SOLiD, Helicos, IonTorrent i PacBio, entre d'altres, ha suposat un impuls important pel que fa a l'obtenció

de gran quantitats de dades en relativament poc temps i a un cost assequible. Degut a l'increment de models de seqüenciadors que cada companyia està desenvolupant i a que cada plataforma de seqüenciació necessita que la preparació de la mostra a ser seqüenciada sigui diferent, existeix abundant bibliografia sobre comparacions de quantitat de dades obtingudes, cost, longitud de lectures, etc (per citar-ne unes poques, Glenn, 2011; Quail *et al.*, 2012, encara que les informacions s'han d'actualitzar contínuament degut a la ràpida evolució d'aquestes plataformes).

Totes aquestes innovacions han fet que el que semblava gairebé impossible, ara es pugui dur a terme. Si com s'ha dit anteriorment, cap a principis-mitjans dels 2000 es comencen a obtenir genomes mitocondrials complets, i encara que prèviament s'havien obtingut fragments de genomes nuclears, Miller *et al.* (2008) obtenen aproximadament el 70% del genoma de mamut (*Mammuthus primigenius*) emprant pèls de dos exemplars, un d'ells d'uns 20.000 anys d'antiguitat. El tipus de material emprat en aquest estudi, pèls, té la particularitat de contenir una elevada proporció d'ADN endogen (Gilbert *et al.*, 2007; 2008). I és aquesta característica dels pèls, juntament amb el seu bon estat de conservació i la tecnologia del moment, la que va permetre seqüenciar el primer genoma complet d'un humà antic, concretament el d'un paleo-esquimal d'uns 4.000 anys trobat a Groenlàndia (Rasmussen *et al.*, 2010). Aquest treball va suposar un estímul important per a posteriors intents d'obtenir material de diferents grups humans, i posava de manifest la importància que podien tenir aquestes tècniques per a la reconstrucció de la història de les poblacions humanes. De fet, des del 2010 s'han anat obtenint genomes complets o quasi complets de diferents grups humans tals com aborígens australians a partir de pèls d'uns 100 anys d'antiguitat (Rasmussen *et al.*, 2011), o de restes humanes de diversos milers d'anys del continent americà (e.g., Raghavan *et al.*, 2013) i europeu (e.g., Keller *et al.*, 2012; Lazaridis *et al.*, 2014).

Respecte a espècies lligades a la línia evolutiva humana, s'ha d'esmentar que ja pels anys 2006-2007 s'havia obtingut una important quantitat d'informació genètica fiable sobre neandertals, *Homo neanderthalensis* (Green *et al.*, 2006; 2008), però l'any 2010 es va poder aconseguir un primer esborrany del genoma de l'espècie (Green *et al.*, 2010). Amb aquests estudis es demostrava que hi ha hagut un flux genètic de neandertals a l'espècie humana moderna, indicant hibridació antiga entre aquestes dues espècies, fet que s'està corroborant en estudis posteriors (e.g., Prüfer *et al.*, 2013). Així, aquest fet suposa que un 1-4% del genomes dels humans no africans deriva dels neandertals.

Però si l'escenari de la paleontologia humana s'està completant gràcies a l'estudi morfològic de noves restes excavades, la disciplina de l'ADN antic també ha proporcionat important sorpreses. La seqüenciació d'unes poques restes d'ossos d'entre 30.000 i 50.000 anys (en concret una falange distal de la mà) procedents de la cova de Denisova a Sibèria va aportar una de les notícies més impactants dels darrers anys (Reich *et al.*, 2010; Meyer *et al.*, 2012). El material genètic obtingut era diferent al dels neandertals i al dels humans moderns i es tractava d'una espècie humana no reconeguda anteriorment.

Un dels factors que s'ha considerat important respecte a la conservació de l'ADN és la seva taxa de degradació post-mortem. Encara que existeixen diversos factors que poden afectar a aquesta taxa, estudis recents suggereixen que obtenir seqüències d'ADN més enllà del milió d'anys és altament improbable (Allentoft *et al.*, 2012). Aquell mateix any, moment en el que no existia cap evidència de que es pogués obtenir ADN de mostres d'uns

centenars de milers d'anys, es va obtenir informació genòmica d'un os polar d'uns 110.000-130.000 anys (Miller *et al.*, 2012), posteriorment, Orlando *et al.* (2013) obtenen el genoma de cavall a partir d'un metàpode trobat a permafrost d'una cronologia al voltant dels 700.000 anys, mentres que Dabney *et al.* (2013) obtenen tot el genoma mitocondrial d'*Ursus deningeri* del Pleistocè Mitjà (d'uns 300.000 anys) fora de permafrost, en concret del jaciment de la Sima de los Huesos a Atapuerca. Del mateix jaciment i amb la mateixa cronologia prové un hominin que comparteix algunes característiques morfològiques amb *Homo heidelbergensis*. Meyer *et al.* (2013) varen obtenir un genoma mitocondrial quasi complet d'aquestes restes i suggereixen que aquest hominin compartiria ancestre amb els humans a la cova de Denisova.

La gran majoria d'estudis en ADN antic s'han basat en l'obtenció de material genètic a partir de mostres d'ossos o dents de vertebrats i, en els pocs casos disponibles, en mostres de pèls. Però dos altres tipus de restes paleontològiques/arqueològiques poden aportar informació sobre paleodietes de tàxons extingits. Primer, cal dir que els copròlits (femtes fossilitzades) han estat emprats per a obtenir ADN dels elements que formaven part de la dieta d'una espècie extingida o inclús per a obtenir ADN de l'espècie diana (e.g., Poinar *et al.*, 1998; 2003; Hofreiter *et al.*, 2003; Campos *et al.*, 2009; Bon *et al.*, 2012; Clack *et al.*, 2012), però també permet la realització d'anàlisis ecològiques o sobre paràsits (e.g., Hofreiter *et al.*, 2003; Wood *et al.*, 2008, 2012, 2013a, 2013b). I segon, els recents estudis genètics de plaques dentals calcificades (també conegut com càlcul dental o tosca) de grups humans antics, estan revolucionant les anàlisis sobre canvis en dieta de poblacions humanes al llarg del temps emprant l'ADN bacterià conservat en aquesta estructura dental (veure la revisió de Weyrich *et al.*, 2014).

Observacions finals

La major part dels avenços metodològics realitzats en biologia molecular durant les últimes dues dècades ha causat un profund impacte en disciplines afins, l'anomenada revolució de les "òmiques": la genòmica, transcriptòmica, la proteòmica, la metabolòmica, filogenòmica, etc. Una evolució similar està tenint lloc en els estudis sobre avaluació de la diversitat biològica, amb la capacitat de seqüenciar tots els genomes presents en una comunitat en particular, que, com s'ha esmentat anteriorment, és ara una tasca incipient en espècies de metazous, la metagenòmica.

Agraïments

Donem les gràcies a Carlos Juan pels seus comentaris i suggeriments que han fet millorar la versió final d'aquest manuscrit. Aquest treball ha estat finançat pel pla Nacional del Govern Espanyol del Ministeri MCINN, projecte CGL2012-33597, parcialment finançat amb fons FEDER de la UE, i per una Marie Curie Internacional Outgoing Fellowship (PB) en el 7è Programa Marc de la Unió Europea (projecte MEDITADNA, PIOF-GA-2011-300854, FP7-PEOPLE).

Genetic and evolutionary tools for the description of biological diversity

The increase in anthropogenic activity in the last two centuries has caused a drastic fragmentation of natural habitats and a decrease of species diversity, in particular in biodiversity hot spots such as the Amazon and Mediterranean basins (Kruess and Tscharntke, 1994; Myers *et al.*, 2000; Brooks *et al.*, 2006; Cardinale *et al.*, 2012). The pace of this destructive process has accelerated so much that many species will be extinct before they can be described as the number of taxonomists is decreasing as well, a process known as taxonomic crisis (Tautz *et al.*, 2003). This issue is pivotal in large systematic groups such as arthropods, in which the number of species known could be half of the estimated values (Odegaard, 2000), and also considering the population levels, since genetic diversities are critical for species fitness and speciation processes (Gugerli *et al.*, 2008).

Three decades ago, evolutionary and systematic studies at the DNA level were only performed in model organisms such as *Drosophila melanogater*, *Caenorhabditis elegans* and *Mus musculus*. Two pioneering methods, however, revolutionized the field, allowing to carry out those surveys in virtually any eukaryotic lineage. The first method made feasible to decipher the sequence of a DNA strand by enzymatic synthesis and use of labeled terminator deoxyribonucleotides (Sanger *et al.*, 1977). The second methodology was the in vitro DNA amplification by Polymerase Chain Reaction (PCR) that uses oligonucleotides (primers), a thermostable DNA polymerase and several cycles composed of three temperature steps to exponentially amplify the number of copies of a particular DNA fragment (Mullis *et al.*, 1986). These technological advances have allowed the reconstruction of phylogenetic trees from DNA sequences, shedding light on the ancestor-descendant relationships from any taxonomic lineage, making possible to compare the retrieved relationships with those based on classical taxonomy and morphological characters (Li, 1997). Until recently, however, most phylogenies were built using a few DNA fragments due to the lack of many conserved 'universal' primers to amplify by PCR the same genes across species. Most 'universal' oligonucleotide primers available are located in the mitochondrial genome (Simon *et al.*, 1994) and in the nuclear ribosomal DNA cluster (Hillis and Dixon, 1991).

Next Generation Sequencing (NGS) methods, fueled by the need to make it cheaper to sequence human genomes, have been developing quickly in last decade leading to a second revolution in molecular phylogenetics. Although nowadays there are several NGS platforms (454, IonTorrent, Solexa and PacificBio), the most widespread sequencing method is the Illumina one due to the high quality and low cost of DNA reads (Quail *et al.*, 2012). For instance, a lane of Illumina HiSeq 2500 produces 330 million reads (sequences) of 150 bp for just € 3,000 in few days of laboratory work. This complex micro-scale method begins by randomly cutting the genome in short DNA fragments ranging from 50 to 300 bp, then adapters are ligated to their ends. Secondly, they are hybridized to a glass microchip coated with complementary oligonucleotide adapters arranged in a geometric pattern. Subsequently, they are enzymatically amplified by a bridge structure to build clusters of identical sequences and hence increasing the signal during the sequencing step. Finally, a snapshot of the entire microchip is captured during each cycle of DNA synthesis, allowing

the detection of light on the clusters that incorporated a particular nucleotide linked to a fluorochrome. The process is explained in detail in this video (http://www.youtube.com/embed/HMyCqWhwB8E?iframe&rel=0&autoplay=1). The NGS methods, among many other applications, have made possible an outstanding improvement in phylogenetic studies producing thousands of genomic sequences from many species (Kawahara *et al.*, 2014; Misof *et al.*, 2014; Pecon-Slattery 2014; Wickett *et al.*, 2014; Zhouq *et al.*, 2014).

The first step in any molecular phylogenetic analysis is to build a multiple alignment of the DNA sequences under study obtained from different species to globally optimize the number of nucleotide identities as an initial homology hypothesis of the nucleotide positions examined (Pons and Vogler, 2006). This process can be fast even if the number of sequences is considerable but the largest datasets can represent a computational challenge during the tree reconstruction step. The algorithms implemented for tree reconstruction were initially based on phenetic distance methods, although they were shown to present several methodological drawbacks making them rarely used nowadays (Hillis *et al.*, 1996). Until recently, most analyses were based on the parsimony criterion as computation algorithms are relatively fast and as is composed of two simple steps. Firstly, a small set of branches of the phylonetic tree (subtree) are pruned and rearranged in another branch of the tree and, then the number of informative nucleotide changes (shared derived characters) are counted on the obtained tree topology (Hillis *et al.*, 1996; Li, 1997). In each iteration, the rearranged topology is accepted and saved into memory only if the number of changes is lower or equal to a previously obtained topology to continue with a new cycle. Parsimony analyses implement an heuristic approach by setting a reasonable number of iterations that start from a random topology. This approach is implemented to avoid getting trapped in a suboptimal solution of the tree space as checking all the possible tree topologies is an unpractical or even, if a large number of sequences are examined, impossible task (Hillis *et al.*, 1996). More recently, the more sophisticated maximum likelihood and bayesian methods have increasingly replaced parsimony in phylogenetic analyses due to the increment of computing power and because they can treat each type of nucleotide substitution independently taking also into account the among-nucleotide site rate variation (Felsenstein, 2004). In short, the maximum likelihood criterion searches the values for the parameters of the evolutionary model that maximize the probability of the observed DNA sequences, while the bayesian method treats parameters as random variables with a prior distribution that are tuned with the observed data (Felsenstein, 2004). In addition, bayesian methods have the advantage that they allow for the evaluation of confidence intervals for each parameter of the evolutionary model implemented, and the statistical fit of different parameter-rich evolutionary models.

Phylogenetic relationships can of course also consider a a temporal framework that would let testing and assessing the congruence among evolutionary transitions and geological events. In the 60s, professors Zuckerkandl and Pauling (1962) introduced the term "molecular clock" showing that mutations on genes under neutral selection, e.g. phylogenetic markers, were relatively constant over evolutionary time allowing to estimate the time separation of two taxa from a common ancestor by counting the number of nucleotide changes across diverging lineages. However, ulterior studies showed that such change constancy is uncommon, so methods that allow a relaxation of the molecular clock

(implementing variation of rates across tree branches) have been developed (Thorne *et al.*, 1998; Sanderson, 2002; Drummond *et al.*, 2006). The improvement of molecular techniques and phylogenetic methods has impacted greatly the assessment of biodiversity at both recent and deep taxonomic and systematic levels, and even on extinct species. These topics are developed in the next sections.

Biodiversity assessment by DNA barcoding

Earlier this century it was proposed the use of the DNA sequence of the region 5' end of the mitochondrial cytochrome oxidase subunit I gene (*cox1*) as an animal species identifier since it contains species-diagnostic nucleotide substitutions (Hebert *et al.*, 2003). The idea is that the species-specificity of the *cox1* sequence could be used to define each species as a particular barcode permits to identify each supermarket product. The success of this species identification based on DNA, together with the awareness of biodiversity and taxonomic crisis, triggered the initiation of the Consortium Barcode of Life (CBOL, English Consortium for the Barcode of Life) to promote between the scientific community the DNA barcoding as a system of recognition and rapid identification of species. This consortium has established protocols to assign accurately the sequence of a new specimen to a taxonomic species or identify it as new undescribed species by comparing morphological taxonomic groups with those clusters of *cox1* sequences outlined by phylogenetic algorithms (Hebert *et al.*, 2003). The gene chosen for animals was *cox1,* but for plants this sequence is inefficient, and instead the genes encoding for the large subunit of Rubisco (*rbcL*) and maturase K (*matK*) have been proposed for plants, while the internal transcriber spacers of nuclear ribosomal RNA (*ITS*) in fungi. Subsequent studies corroborated the many advantages of DNA barcodes: universality and homology of DNA sequences opposed to the, in many cases, ambiguity and lack of species-specificity of morphological characters. In addition, DNA barcodes can incorporate information at the systematic level, can be easily automated at a relatively low-cost in different contexts even for non-specialists, the information they provide is highly portable and of immediate access, and finally the methods can be useful in a wide range of phylogenetic and taxonomic levels (Hebert *et al.*, 2003; Hajibabaei *et al.*, 2005). However, DNA barcodes have also drawbacks: the set up an optimal cut-off value of genetic divergence to delimit species boundaries may be uncertain as the rate of nucleotide substitution is highly variable across lineages, the lack of genetic divergence may not allow to discriminate closely related species, genetic clusters defined based on the evolutionary history of a single gene can not unequivocally be extrapolated into species boundaries, and finally and most importantly, the taxonomic knowledge built since Linnaeus over the past three centuries can not be easily incorporated in the new framework (Pons *et al.*, 2006; Rubinoff *et al.*, 2006).

The initial success of many DNA barcoding projects made that several american and international institutions joined efforts to promote large-scale projects to obtain thousands of DNA barcode sequences from complete taxonomic groups or from particular geographic regions: FISH-BOL (all fish species on a global scale), ABB Initiative (birds), plant DNA Barcode Project, fungal Barcode, INBIPS (invasive species and parasites), and BioCode Moorea (all macroscopic species from Moorea). As is in January 2015, the website of the

University of Guelph (http://www.barcodinglife.org/) has publicly available 3676066 DNA barcodes mostly from specimens collected in developed countries (Europe, North America and Australia). For instance, there are deposited 29,992 DNA barcodes of birds from 192 countries that account for 4,778 genetic clusters or putative species with 4,080 taxonomically identified, 61,846 of mammals and from 164 countries (3,224 / 2,108), 2,380,002 of insects from 216 countries (80,392 / 260,676), 79,829 of molluscs from 156 countries (12,658 / 8,840), and 4,740 of flatworms from 85 countries (568 / 670). The analysis of this huge data suggest that most of the putative (described + cryptic + unknown) species delimited by DNA barcoding in birds and mammals were already described species by classical taxonomy, but on the other hand, that the DNA barcoding approach can identify previously unkown species, particularly in molluscs and clearly insects. In addition, the results in flatworms suggest the many cases of misidentification based on morphology.

DNA barcoding has been successfully applied to different fields such as detecting food fraud in processed or served food by determining the composition of animal and plant species (Lowenstein *et al*., 2009; Di Pinto *et al*., 2013; Vartak *et al*., 2015), to ascertain the plant species used by bees to make honey (Bruni *et al*., 2015), characterizing the diet from stomach contents and fecal samples of animal species (Pons, 2006; Jurado-Rivera *et al*., 2009; Ibanez *et al*., 2013), identifying illegal trade with CITES-listed species (Liu *et al*., 2013), detection of parasite eggs and larvae (Gariepy *et al*., 2014), or finding out which species are bitten by hematophagous insects (Pettersson *et al*., 2013). The application of current NGS technologies to DNA barcoding projects could led to scale studies from hundreds of specimens to millions of individuals from a complete community particularly those of very small size (McMahon *et al*., 2014).

Metazoan phylogenomics

The origin of multicellular organisms has captivated the attention of evolutionary biologists for many years. Two recent studies tried to solve the question of whether multicellurarity appeared just once or evolved several times independently by using an approach termed transcriptomics (Dunn *et al*., 2008; Ruiz-Trillo *et al*., 2008). In transcriptomics, isolated messenger RNAs (representing genes that are being expressed in a particular tissue or individual) are converted to cDNAs by means of a reverse transcriptase, and later amplified by PCR. These cDNAs are finally sequenced by Sanger method, or nowadays, by NGS techniques. One of the studies focused on discover which extant unicellular eukaryotic taxon or taxa are sister to all Metazoa (Ruiz-Trillo *et al*., 2008), To achieve that, the authors generated the sequences for 110 nuclear and 13 mitochondrial genes for 12 species including sponges, annelids, insects, and vertebrates with a total nuclear DNA sequences of 63,000 bp and 7,800 bp from mitochondrial genes. The phylogenetic analyses based on maximum likelihood and bayesian criteria were performed at the protein level to avoid saturation from multiple nucleotide substitutions, reducing the dataset to 21,000 nuclear genes aminoacids and 2,600 mitochondrial aminoacids (Fig. 1). This study also included four protistan taxa that are considered to be closely related to Metazoa (the choanoflagellate *Monosiga ovata*; the ichthyosporeans *Sphaeroforma arctica*

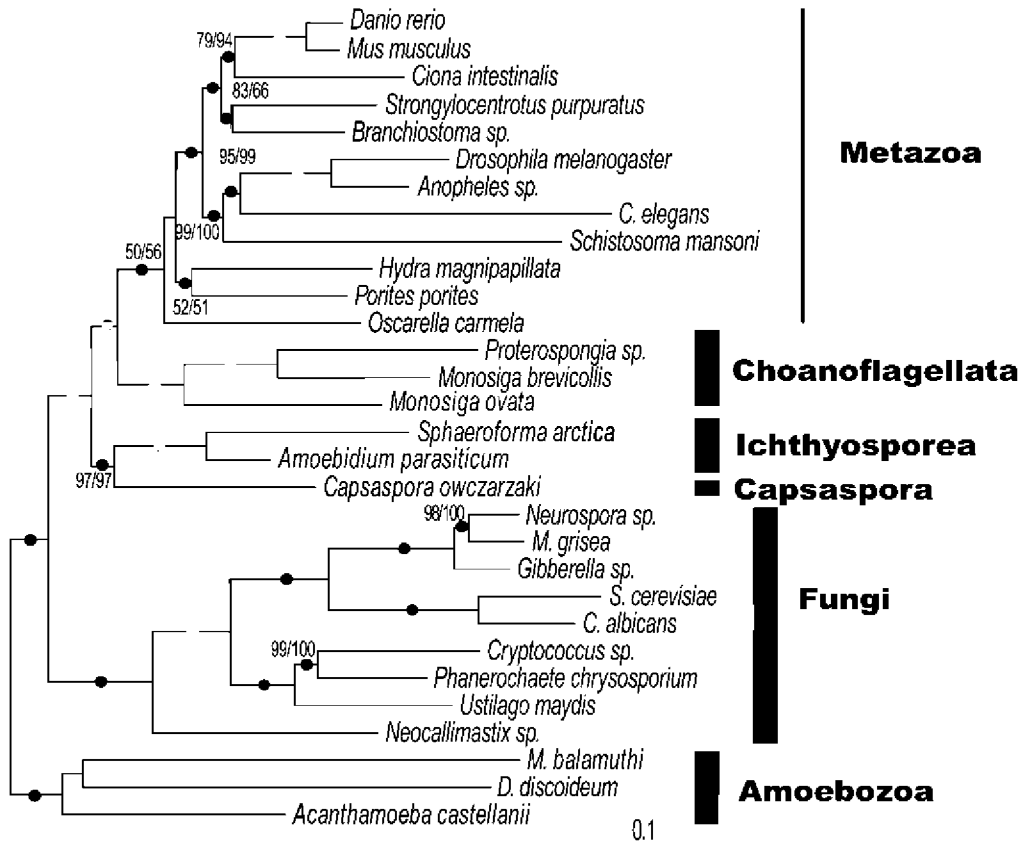


Fig. 1. Phylogenetic tree of the Opisthokonta based the Bayesian analysis 110 nuclear proteins sequences. Black dots on nodes indicate posterior probability values of 1.0, and grey ones if maximum likelihood bootstrap analysis also yields 100% support. Modified from Ruiz-Trillo *et al.*, 2008.

Fig. 1. *Arbre filogenètic dels Opisthokonta basat en l'anàlisi baiesiana de 110 seqüències de proteïnes nuclears. Els punts negres en els nodes indiquen els valors de probabilitat posterior de 1.0, i els grisos si els valors de suport de bootstrap de màxima versemblança són de també del 100%. Modificat de Ruiz-Trillo* et al. *(2008).*

and *Amoebidium parasiticum* and the amoeba *Capsaspora owczarzaki)* and two distant Amoebozoa outgroups (*Acanthamoeba castellanii* and *Mastigamoeba balamuthi*). The nuclear and mitochondrial phylogenetic trees were congruent, indicating that the closest group to Metazoa are Choanoflagellates followed by Capsaspora and Ichthyosporea lineages. Interestingly, primitive unicellular groups such as Capsaspora already present in their genome genes homologous to the gene membrane-associated guanylate kinase (MAGI) coding for a protein that has an important role in cell communication.

The second study was focused to resolve the phylogenetic relationships among the major taxonomic lineages and body plans within Metazoa (Dunn *et al.*, 2008). The dataset

obtained in this research was composed of 150 aminoacid sequences from 71 metazoan taxa plus six outgroup species that were obtained by a transcriptomic approach as explained above (Fig. 2). This project generated nearly 40,000,000 bp (40 Mb) of DNA sequence data. The phylogenetic results fully corroborated the monophyly of several taxa previously defined by classical systematics such Metazoa, Deuterostomia (Chordata, Echinodermata + Hemichordata) and Protostomia. Nonetheless, the latter included new groups incongruent with classical systematics such as Ecdysozoa (Pancrustacea+Nematoda) which common feature is having the body covered by a cuticle, and Lophotrochozoa that comprises specimens with a characteristic feeding organ called lophophore with a ring of ciliated tentacles surrounding the mouth (Phoronida and Brachiopoda) or species with trochophore larvae showing two characteristic middle bands of cilia (Mollusca, Anelida, Echiura, Sipuncula and Nemertea). Other interesting findings were that Celomata and Mandibulata are not monophyletic and, therefore, their systematic classification should revised, that Ctenophora is basal to all other metazoans instead of Cnidaria and Porifera, and finally that Choanoflagellata was the sister group to Metazoa thus corroborating Ruiz-Trillo *et al.* (2008) findings. Similar phylogenomic approaches based on transcriptomics have been implemented to resolve phylogenetic relationships and other evolutionary questions among insects (Misof *et al.*, 2014), land plants (Wickett *et al.*, 2014), mammals (Zhouq *et al.*, 2014), butterflies and moths (Kawahara and Breinholt, 2014) and primates (Pecon-Slattery, 2014).

A phylogenomic approach to the diversification of the global avian fauna

Birds have always have had the attention of naturalists due to their remarkable characteristics such flying capacity, vivid and complex plumage patterns, singing abilities, and intricate courtships plus complex reproductive and social behaviors. This animal group has been studied extensively by biologists, trying to shed light on the evolution of several traits but also to understand their diversificacion through time, contrasting the results with their paleontological record (Brown *et al.*, 2008; Pacheco *et al.*, 2011). For instance, it is not clear if radiation of neoavian birds (e.g., songbirds, parrots, pigeons, and others) was gradual or explosive (Pacheco *et al.*, 2011). Previous studies failed give clear answers to some of these questions as molecular phylogenetic trees were not fully resolved and relationships supported as DNA sequence datasets was not enough informative. Jarvis *et al.* (2014) have recently tackled this problem by sequencing the complete genomes of 48 bird species representing the major taxonomic lineages. The first step was to select the homologous sequences on each genome to construct the final dataset that included 40% of the proteome (genes coding for proteins, although they represent about 3.5% of the total genomic DNA sequences of a bird), intronic (non-coding gene sequences) regions from 2,516 genes, and 3,756 non-coding ultraconserved DNA elements, including 1000 bp of flanking sequences. The total dataset comprised the staggering amount of about 42 millions base pairs of DNA sequences (42 Mb). Dating the nodes of the bird phylogeny required the transformation of the number of nucleotide substitutions to an estimation of millions of years by constraining the age of 19 nodes in the phylogenetic tree to the age of particular fossil birds as a minimum value, plus setting the age for some external groups and a

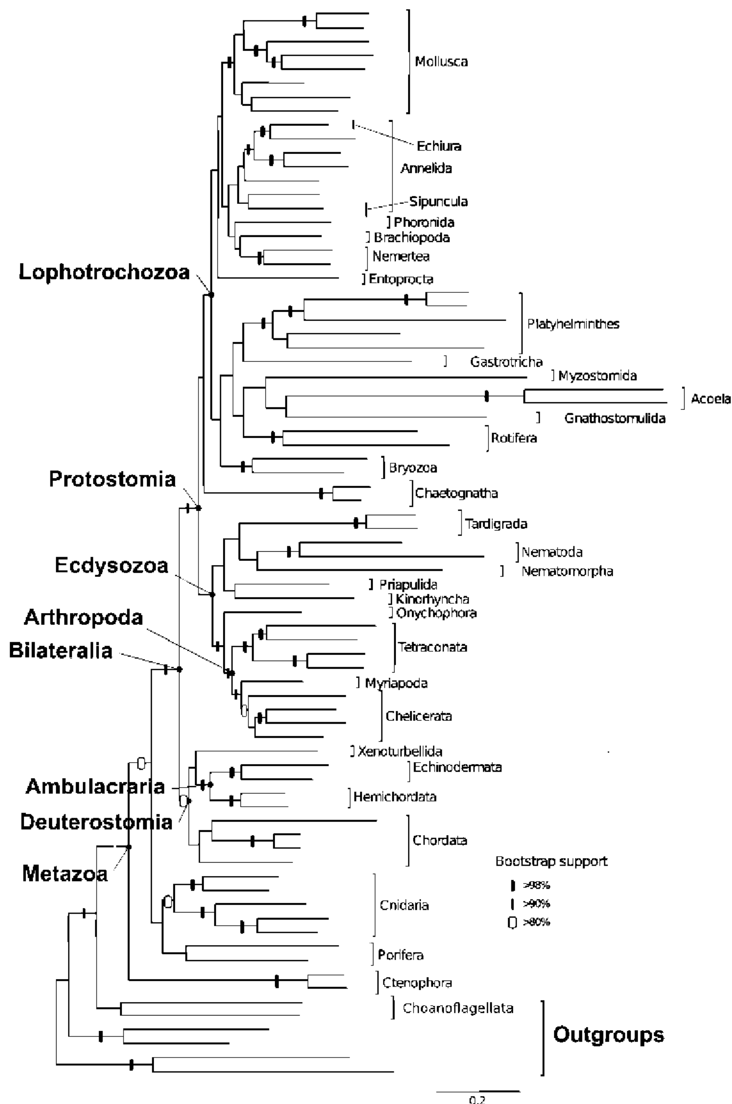


Fig. 2. Phylogenetic tree of 71 metazoan species, including most important taxonomic lineages, plus 6 outgroups based on maximum likelihood analysis of 42 Mb of protein sequences. Branches with different symbols indicates bootstrap support values ranging from 80 to 100%. Modified from Dunn *et al.*, 2008

Fig. 2*. Arbre filogenètic de 71 espècies de metazous, incloent membres dels llinatges taxonòmics més importants, a més de 6 grups externs basats en l'anàlisi de màxima versemblança de 42 Mb de seqüències de proteïnes. Branques amb diferents símbols indiquen els valors de suport de bootstrap que van des de 80 a 100%. Modificat de Dunn* et al. *(2008).*

.
minimum age for modern birds (Neornithes) to 99.6 Ma. This phylogenomic analysis comprised a so large dataset that required nine super-computers to get the job done in few weeks, with a computing time in an average PC equivalent to 400 years! The results suggest that the two major lineages of birds, Palaeognathae and Neognathae, split about 100 Ma in the Late Cretaceous, and Passeriformes, which represent 60% of the 10,400 avian species, radiated around 39 Ma. This robust phylogeny also allowed studying the evolution of some avian novel traits. For instance, language in birds evolved independently several times, and birds of prey are polyphiletic since they are basal to two independent radiations, Australaves and Afroaves. This study also showed that some taxonomic groups defined by shared characters or behaviors are actually not monophyletic, and therefore those features evolved several times and constitute evolutionary convergences. This include for example footpropelled diving trait of grebes (Columbea) and cormorants (Passerea), and wading-feeding trait of flamingos (Columbea), ibises and egrets (Passerea). Finally, one of the most interesting results of this study is that confirm that the radiation of most extant species (Neonaves) occurred explosively in 10 - 15 Ma coincident with the Cretaceous to Paleogene (K-Pg) transition, this is coeval with the extinction of their sister dinosaur group during the Terciary that produced empty ecological niches.

Extinct species and Ancient DNA

Molecular genetics has undergone a steady period of expansion in the last 30 years from which took advantage the ancient DNA, i.e. DNA sequences from fossil or sub-fossil samples. The development of this discipline has been extraordinarily fast, not only with regard to techniques and laboratory methods but also regarding the topics to be addressed. At the beginning, the primary goal was to recover DNA from extinct species to shed light on the phylogenetic relationships among extinct and extant species though the complexity of the questions that can be currently solved has increased considerably: population studies over time, whole genome sequencing, etc.

Since the initial recovery of ancient DNA in early 80's from a quagga, an extinct horse family (Higuchi *et al.*, 1984), and from mummified human remains (Pääbo, 1985), many researchers have published remarkable results on ancient molecular genetics. After these first studies, ancient DNA analysis took advantage of the advances of PCR techniques. Early studies focused on obtaining short mitochondrial DNA fragments for phylogenetic studies due to DNA molecules degrades in shorter fragments over time and because eukaryotic cells contain many mitochondria (there are between 100 and 10,000 copies of mitochondrial genomes, which have not recombination). Among others, the studies about the moas from New Zealand (Cooper *et al.*, 1992) and thylacine from Tasmania (Thomas *et al.*, 1989) are adroit examples. It is worth mentioning that first results already pinpointed the importance of these findings for other research fields such as paleontology, archeology, anthropology, or even for the species conservation. On the other hand, the early steps of this discipline were not immune to difficulties such as contamination by exogenous modern DNA which is the main source of error. Thus, studies that recovered DNA from extremely old samples such as dinosaurs bones (Woodward *et al.*, 1994; An *et al.*, 1995; Li *et al.*,

1995), Miocene plant remains (Golenberg *et al.*, 1990; Soltis *et al.*, 1992) or insects embedded in amber (Cano *et al.*, 1993) were subsequently questioned and rejected due to this issue (see, for example, Hedges and Schweitzer, 1995; Zischler *et al.*, 1995; Ward *et al.*, 1997; Penney *et al.*, 2013). In this regard, it has been a continuous progress in establishing strict protocols, mechanisms and control guidelines to prevent contamination in the extraction methods such as replication of the procedures in two or more independent laboratories to ensure the accuracy and reliability of results (e.g. Cooper and Poinar, 2000). In fact, most of the problems concomitant to ancient DNA have been understood during the development of new studies, e.g. DNA fragmentation and C to T and G to A misincorporations that are common in ancient sequences caused by cytosine deamination. Interestingly, these type of nucleotide transitions are now important to verify that the sequences obtained originate from genuine ancient DNA (eg, Briggs *et al.*, 2007). Thus, all methodological advances on DNA extraction, cloning, genetic library construction and enrichment by hybridization with RNA probes, along with the development of NGS platforms led to the attainment of complete mitochondrial genomes of several animal species as the New Zealand moas (Cooper *et al.*, 2001; Haddrath and Baker, 2001), the woolly mammoth *Mammuthus primigenius* (Krause *et al.*, 2006; Rogaev *et al.*, 2006), the cave bear *Ursus spelaeus* (Bon *et al.*, 2008; Krause *et al.*, 2008), the woolly rhinoceros *Coelodonta antiquitatis* (Willerslev *et al.*, 2009) or the European cow *Bos primigenius* (Edwards *et al.*, 2010).

About ten years ago it was unthinkable to sequence the complete genome of extinct species or taxonomic groups, especially from ancient human lineages, and the financial cost and manual handling required were surpassing. An important issue, among others, was that each experiment required a significant amount of starting material that was often unobtainable for most fossil samples. This was particularly disadvantageous in projects working on nuclear genes as most of them are in low copy numbers in the genome. To overcome this issue, targeted DNA has to be amplified by a series of PCRs, amplicons cloned using bacterial vectors, and then hundreds or thousands of clones sequenced individually by Sanger method. However, the development of the emulsion PCR and the sequencing platform 454 of Life Sciences (later acquired by Roche) in 2005 allowed to amplify each DNA molecule in an emulsion oil-water droplet and obtaining thousands of individual sequencing reactions in parallel that was a faster and more effective method than classical cloning. The ulterior development of other NGS such as Solexa (later acquired by Illumina), SOLiD, Helicos, IonTorrent and PacBio has been a major boost in terms of obtaining large amounts of sequence data in relatively short delivery time and affordable cost. The technology behind those sequencing platforms is so varied that each company requires a specific protocol of sample preparation before it can be sequenced. There is an abundant literature that compares amount of data delivered, read length, cost, etc (e.g., Glenn, 2011; Quail *et al.*, 2012) but this information have to be updated continuously due to the rapid evolution of these platforms.

All these innovations made feasible some projects that were impracticable few years ago. As mentioned above, some projects have already sequenced some nuclear fragments and even complete mitochondrial genomes in mid 2000's. But just few years later, Miller and colleagues made a step forward by sequencing approximately 70% of the genome of the woolly mammoth (*Mammuthus primigenius*) using hair from two specimens, one of

them about 20,000 years old (Miller *et al.*, 2008). The success of the project was partly due to that hair has the distinction of containing a high proportion of endogenous DNA (Gilbert *et al.*, 2007, 2008). This feature of hair, together with the good preservation of the sample and the NGS technology of the time, enabled to secure the first complete human genome from old remains, a paleo-Eskimo who lived in Greenland about 4,000 years ago (Rasmussen *et al.*, 2010). This work was an important impulse for posterior efforts to recover material from other human lineages, and also acknowledged that these molecular techniques could be crucial to reconstruct the history of human populations. In fact, complete or nearly complete genomes of different human groups have been reported since 2010 such as Aboriginal Australians from hair about 100 years old (Rasmussen *et al.*, 2011), or human remains several thousands of years old from the American (eg, Raghavan *et al.*, 2013), and European (e.g., Keller *et al.*, 2012; Lazaridis *et al.*, 2014) continents.

In 2006-2007, some projects working on lineages closely related to modern humans had already acquired a significant amount of reliable genetic information about Neanderthals, *Homo neanderthalensis* (Green *et al.*, 2006, 2008), but years later it was possible to get the first draft of the genome of a Neanderthal (Green *et al.*, 2010). These studies demonstrated that existed genetic flow between Neanderthals and modern humans, i.e. an ancient hybridization between these two species, that was corroborated elsewhere (e.g. Prüfer *et al.*, 2013). Thus, these results suggest that about of 1-4% of the genome of the current non-African human populations is derived from Neanderthals.

While the scenario of human paleontology is being accomplished thanks to new excavated remains, the discipline of ancient DNA has also provided astonishing findings. For instance, the sequencing of a few bone remains with a chronology between 30,000 and 50,000 years (in particular a hand distal phalanx) from the Denisova Cave in Siberia brought one of the most shocking news in recent years (Reich *et al.*, 2010; Meyer *et al.*, 2012). The genetic material obtained was different from sequences of Neanderthals and modern humans, i.e. there was a new human species previously unknown.

One of the important factors related to DNA preservation is its rate of postmortem degradation. Although there are several factors that can affect this rate, recent studies suggest that obtaining DNA sequences beyond one million years old is extremely unlikely (Allentoft *et al.*, 2012). However, few moths later, this statement was challenged since one study obtained genomic sequences of a polar bear from about 110,000 to 130,000 years old (Miller *et al.*, 2012), and, later on, the genome of a horse from a metapod found in permafrost with a chronology of around 700,000 years (Orlando *et al.*, 2013). In addition, Dabney *et al.* (2013) recovered the entire mitochondrial genome of the bear *Ursus deningeri* from Middle Pleistocene (about 300,000 years ago) out of permafrost, specifically from the site of Sima de los Huesos in Atapuerca. From the same site and chronology come a hominin that shares some morphological characteristics with *Homo heidelbergensis*. Meyer *et al.* (2013) obtained an almost complete mitochondrial genome of this hominin and genetics analysis suggest that these remains share a common ancestor with humans from Denisova Cave.

The vast majority of ancient DNA studies have been based on genetic molecules isolated from vertebrate bones and teeth, except few cases using hair samples, but two other types of paleontological / archaeological remains can supply information about paleo-diet of extinct taxa. First of all, coprolites (fossilized feces) were used to isolate DNA from the

components of the diet of extinct species or even to obtain DNA from the specimen itself (e.g. Poinar *et al.*, 1998, 2003; Hofreiter *et al.*, 2003; Campos *et al.*, 2009; Bon *et al.*, 2012; Clack *et al.*, 2012), but also to carry out ecological studies or about parasite content (eg, Hofreiter *et al.*, 2003; Wood *et al.*, 2008, 2012, 2013, 2013b). Secondly, recent genetic studies of calcified dental plates (also known as dental calculus) of ancient human groups, are revolutionizing the analysis of changes in diet of human populations over time using bacterial DNA preserved in this dental structure (see review of Weyrich *et al.*, 2014).

Final remarks

Most of the methodological advances made in molecular biology during the last two decades caused a deep impact in related disciplines, so-called the 'omics' revolution: genomics, transcriptomics, proteomics, metabolomics, phylogenomics, etc. Similar evolution is taking place on studies about biological diversity assessment making able to sequence all genomes present in a particular community which, as mentioned above, is now an ongoing task on metazoan species, metagenomics.

Acknowledgements

We thank Carlos Juan for this comments and suggestions which considerably improved the final version of the manuscript. This work was supported by Spanish MCINN grant CGL2012-33597 partially financed with EU FEDER funds, and by a Marie Curie International Outgoing Fellowship (PB) within the 7th European Community Framework Programme (project MEDITADNA, PIOF-GA-2011-300854, FP7-PEOPLE).

References

Allentoft, M.E., Collins, M., Harker, D., *et al.* 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B*, 279: 4724-4733.

An, C.-C., Li, Y., Zhu, Y.-X., *et al.* 1995. Molecular cloning and sequencing of the 18S rDNA from specialized dinosaur egg fossil found in Xixia Henan, China. *Acta Sci. Nat. Univ. Pekinensis*, 31: 140-147.

Bon, C., Caudy, N., de Dieuleveult, M., *et al.* 2008. Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proc. Natl. Acad. Sci. USA*, 105: 17447-17452.

Bon, C., Berthonaud, V., Maksud, F., *et al.* 2012. Coprolites as a source of information on the genome and diet of the cave hyena. *Proc. R. Soc. Lond. B*, 279: 2825-2830.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA*, 104: 14616-14621.

Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., *et al.* 2006. Global Biodiversity Conservation Priorities. *Science*, 313: 58-61.

Brown, J.W., Rest, J.S., García-Moreno, J., *et al.* 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol.*, 6: 6.

Bruni, I., Galimberti, A., Caridi, L., *et al.* 2015. A DNA barcoding approach to identify plant species in multiflower honey. *Food. Chem.*, 170: 308-315.
Campos, P.F., Willerslev, E., Mead, J.I., *et al.* 2009. Molecular identification of the extinct mountain goat, *Oreamnos harringtoni* (Bovidae). *Boreas*, 39: 18-23.
Cano, R.J., Poinar, H.N., Pieniazek, N.J., *et al.* 1993. Amplification and sequencing of DNA from a 120-135 million-years-old weevil. *Nature* 363: 536-538.
Cardinale, B.J., Duffy, J.E., Gonzalez, A., *et al.* 2012. Biodiversity loss and its impact on humanity. *Nature*, 486: 59-67.
Clack, A.A., MacPhee, R.D.E. i Poinar, H.N. 2012. *Mylodon darwinii* DNA sequences from ancient fecal hair shafts. *Annals of Anatomy*, 194: 26-30.
Cooper, A. i Poinar, H.N. 2000. Ancient DNA: do it right ot not at all. *Science*, 289: 1139.
Cooper, A., Mourer-Chauviré, C., Chambers, G.K., *et al.* 1992. Independent origins of New Zealand moas and kiwis. *Proc. Natl. Acad. Sci. USA*, 89: 8741-8744.
Cooper, A., Lalueza-Fox, C., Anderson, S., *et al.* 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, 409: 704-707.
Dabney, J., Knapp, M., Glocke, I., *et al.* 2013. Complete mitochondrial genome structure of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA*, 110: 15758-15763.
Di Pinto, A., Di Pinto, P., Terio, V., *et al.* 2013. DNA barcoding for detecting market substitution in salted cod fillets and battered cod chunks. *Food. Chem.*, 141:1757-1762.
Drummond, A.J., Ho, S.Y.W., Phillips, M.J., *et al.* 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.*, 4: e88.
Dunn, C.W., Hejnol, A., Matus, D.Q., *et al.* 2008 Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452: 745-749.
Edwards, C.J., Magee, D.A., Park, S.D., *et al.* 2010. A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS ONE*, 5: e9255.
Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer Associates Inc. Publishers, Sunderland. 580 pp.
Gariepy, T.D., Haye, T. i Zhang, J. 2014. A molecular diagnostic tool for the preliminary assessment of host-parasitoid associations in biological control programmes for a new invasive pest. *Mol. Ecol.*, 23: 3912-3924.
Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., *et al.* 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*, 317: 1927-1930.
Gilbert, M.T.P., Drautz, D.I., Lesk, A.M., *et al.* 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl Acad. Sci. USA*, 105: 8327-8332.
Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. *Mol. Ecol. Res.*, 11: 759-769.
Golenberg, E.M., Giannasi, D.E., Clegg, M.T., *et al.* 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature*, 344: 656-658.
Green, R.E., Krause, J., Ptak, S.E., *et al.* 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444: 330-336.
Green, R.E., Malaspinas, A.-S., Krause, J., *et al.* 2008. A complete Neandertal mitochondrial genome sequence determined. *Cell*, 134: 416-426.
Green, R.E., Krause, J., Briggs, A.W., *et al.* 2010. A draft sequence of the Neandertal genome. *Science*, 328: 710-722.
Gugerli, F., Englisch, T., Niklfeld, H., *et al.* 2008. Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspect. Plant Ecol. Evol. Syst.*, 10: 259-281.
Haddrath, O. i Baker, A.J. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. R. Soc. Lond. B*, 268: 939-945.

Hajibabaei, M., DeWaard, J.R., Ivanova, N.V., *et al.* 2005. Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. Lond. B*, 360: 1959-1967.

Hebert, P.D.N., Cywinska, A. i Ball, S.L. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B*, 270: 313-321.

Hedges, S.B. i Schweitzer, M.H. 1995. Detecting dinosaur DNA. *Science*, 268: 1191-1192.

Higuchi, R., Bowman, B., Freiberger, M., *et al.* 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312: 282-284.

Hofreiter, M., Betancourt, J.L., Pelliza-Sbriller, A., *et al.* 2003. Phylogeny, diet, and habitat o an extinct ground sloth from Cuchillo Curá, Neuquén Province, southwest Argentina. *Quat. Res.*, 59: 364-378.

Hillis, D.M. i Dixon, M.T. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quat. Rev. Biol.*, 664.: 411-453.

Hillis, D.M., Moritz, C. i Mable B.K. 1996. *Molecular Systematics*. Sinauer Associates, Inc. Publishers, Sunderland. 655 pp.

Ibanez, S., Manneville, O., Miquel, C., *et al.* 2013. Plant functional traits reveal the relative contribution of habitat and food preferences to the diet of grasshoppers. *Oecologia*, 173: 1459-1470.

Jarvis, E.D., Mirarab, S., Aberer, A.J., *et al.* 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 12: 1320-1331.

Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., *et al.* 2009. DNA barcoding insect-host plant associations. *Proc. R. Soc. Lond. B.*, 276: 639-648.

Kawahara, A.Y. i Breinholt, J.W. 2014. Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proc. R. Soc. Lond. B*, 281: 20140970.

Keller, A., Graefen, A., Ball, M., *et al.* 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.*, 3: 698.

Krause, J., Dear, P.H., Pollack, J.L., *et al.* 2006. Multiplex amplification of the mammoth mitochondrial genome and evolution of Elephantidae. *Nature*, 439: 724-727.

Krause,J., Unger,T., Nocon, A., *et al.* 2008 Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.*, 8: 220.

Kruess A. i Tscharntke, T. 1994. Habitat fragmentation, species loss, and biological control. *Science*, 264: 1581-1584.

Lazaridis, I., Patterson, N., Mittnik, A., *et al.* 2014.Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513: 409-413.

Li, W.H. 1997. *Molecular Evolution*. Sinauer Associates Inc. Publishers, Sunderland. 487 pp.

Li, Y., An,C.-C., Zhu,Y.-X., *et al.* 1995. DNA isolation and sequence analysis of dinosaur DNA from Cretaceous dinosaur egg in Xixia Henan, China. *Acta Sci. Nat. Univ. Pekinensis*, 31: 148-152.

Liu, S.-Y.V., Chan, C.-L.C., Lin, O., *et al.* 2013. DNA barcoding of shark meats identify species composition and CITES-listed species from the markets in Taiwan. *PLoS ONE*, 8: e79373.

Lowenstein, J.H., Amato, G. i Kolokotronis, S. 2009. The real maccoyii: identifying tuna sushi with DNA barcodes - Contrasting characteristic attributes and genetic distances. *PLoS ONE*, 4: e7866.

McMahon, B.J., Teeling, E.C. i Höglund, J. 2014. How and why should we implement genomics into conservation? *Evol. Appl.*, 7: 999-1007.

Meyer, M., Kircher, M., Gansauge, M.-T., *et al.* 2012. A high-coverage genome sequence rom an Archaic Denisovan individual. *Science*, 338: 222-226.

Meyer, M., Fu, Q., Aximu-Petri, A., *et al.* 2013. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, 505: 403-406.

Miller, W., Drautz, D.I., Ratan, A., *et al.* 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, 456: 387-390.

Miller, W., Schuster, S.C., Welch, A.J., *et al.* 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl Acad. Sci. USA*, 109: E2382-E2390.
Misof, B., Liu, S., Meusemann, K., *et al.* 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346: 763-767.
Mullis, K., Faloona, F., Scharf, S., *et al.* 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, 51: 263-273.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., *et al.* 2000. Biodiversity hotspots for conservation priorities *Nature*, 403: 853-858.
Odegaard, F. 2000. How many species of arthropods? Erwin's estimate revised. *Biol. J. Linn. Soc.*, 71: 583-597.
Orlando, L., Ginolhac, A., Zhang, G., *et al.* 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499: 74-78.
Pääbo, S. 1985. Molecular cloning of ancient Egyptian mummy DNA. *Nature*, 314: 644-645.
Pacheco, M.A., Battistuzzi, F.U., Lentino, M., *et al.* 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol. Biol. Evol.*, 28: 1927-1942.
Pecon-Slattery, J. 2014. Recent advances in primate phylogenomics. *Annu. Rev. Anim. Biosci.*, 2: 41-63.
Penney, D., Wadsworth, C., Fox, G., *et al.* 2013. Absence of ancient DNA in sub-fossil insect inclusions preserved in 'Anthropocene' Colombian Copal. *PLoS ONE*, 8: e73150.
Pettersson, E., Bensch, S., Ander, M., *et al.* 2013. Molecular identification of bloodmeals and species composition in Culicoides biting midges. *Med. Vet. Entomol.*, 27: 104-112.
Poinar, H.N., Hofreiter, M., Spaulding, W.G., *et al.* 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science*, 281: 402-406.
Poinar, H.N., Kuch, M., McDonald, G., *et al.* 2003. Nuclear gene sequences from a Late Pleistocene sloth coprolite. *Curr. Biol.*, 13: 1150-1152.
Pons, J. 2006. DNA-based analysis for identifying preys from total DNA extracted from whole predators specimens using non-destructive methods, and arthropod 'universal' primers. *Mol. Ecol. Notes*, 6: 623-626.
Pons, J. i Vogler, A.P. 2006 Size, frequency, and phylogenetic signal of multiple-residue indels in sequence alignment of introns. *Cladistics*, 22: 144-156.
Pons, J., Barraclough T.G., Gómez-Zurita, J., *et al.* 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.*, 55: 595-609.
Prüfer, K., Racimo, F., Patterson, N., *et al.* 2013. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505: 43-49.
Quail, M.A., Smith, M., Coupland, P., *et al.* 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13: 341.
Raghavan, M., Skoglund, P., Graf, K.E., *et al.* 2013. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, 505: 87-91.
Rasmussen, M., Li, Y., Lindgreen, S., *et al.* 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463: 757-762.
Rasmussen, M., Guo, X., Wang, Y., *et al.* 2011. An Aboriginal Australian genome reveals separate human dispersal into Asia. *Science*, 334: 94-98.
Reich, D., Green, R.E., Kircher, M., *et al.* 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468: 1053-1060.
Rogaev, E.I., Moliaka, Y.K., Malyarchuk, B.A., *et al.* 2006. Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biol.*, 4: e73.
Rubinoff, D., Cameron, S. i Will, K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *J. Hered.*, 97: 581-594.

Ruiz-Trillo, I., Roger, A.J., Burger, G., *et al.* 2008. A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.*, 25: 664-672.

Sanderson, M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.*, 19: 101-109.

Sanger, F., Nicklen, S. i Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.*, 74: 5463-5467.

Simon, C., Frati, F., Beckenbach, A., *et al.* 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Ent. Soc. Am.*, 87: 651-701.

Soltis, P.S., Soltis, D.E. i Smiley, C.J. 1992. An rbcL sequence from a Miocene *Taxodium* (bald cypress). *Proc. Natl. Acad. Sci. USA*, 89: 449-451.

Tautz, D., Arctander, P., Minelli, A., *et al.* 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.*, 182: 70-74.

Thomas, R.H., Schaffner, W. i Wilson, A.C. 1989. DNA phylogeny of the extinct marsupial wolf. *Nature*, 340: 465-467.

Thorne, J.L., Kishino, H. i Painter, I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.*, 15: 1647-1657.

Vartak, V.R., Narasimmalu, R., Annam, P.K., *et al.* 2015. DNA barcoding detected improper labelling and supersession of crab food served by restaurants in India. *J. Sci. Food Agric.*, 95: 359-366.

Ward, H.-L., Yan, Z.-Y. i Jin, D.-Y. 1997. Reanalysis of published DNA sequence amplified from Cretaceous dinosaur egg fossil. *Mol. Biol. Evol.*, 14: 589-591.

Weyrich, L.S., Dobney, K. i Cooper, A. 2014. Ancient DNA analysis of dental calculus. *J. Hum. Evol.*, 79: 119-124.

Wickett, N.J., Mirarab, S., Nguyen, N., *et al.* 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA*, 111: E4859–E4868.

Willerslev, E., Gilbert, M.T.P., Binladen, J., *et al.* 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evol. Biol.*, 9: 95.

Wood, J.R., Rawlence, N.J., Rogers, G.M., *et al.* 2008. Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quat. Sci. Rev.*, 27: 2593-2602.

Wood, J.R., Wilmshurst, J.M., Wagstaff, S.J., *et al.* 2012. High resolution coproecology: using coprolites to reconstruct the habits and habitats of New Zealand's extinct upland moa (*Megalapteryx didinus*). *PLoS ONE*, 7: e40025.

Wood, J.R., Wilmshurst, J.M., Rawlence, N.J., *et al.* 2013a. A megafauna's microfauna: gastrointestinal parasites of New Zealand's extinc moa (Aves: Dinornithiformes). *PLoS ONE*, 8: e57315.

Wood, J.R., Wilmshurst, J.M., Richardson, S.J., *et al.* 2013b. Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes). *Proc. Natl Acad. Sci. USA*, 110: 16910-16915.

Woodward, S.R., Weyand, N.J. i Bunnell, M. 1994. DNA sequence from Cretaceous period bone fragments. *Science*, 266: 1229-1232.

Zhouq, X., Sun, F., Xu, S., *et al.* 2014. The position of tree shrews in the mammalian tree: comparing multi-gene analyses with phylogenomic results leaves monophyly of Euarchonta doubtful. *Int. Zool.*: doi:10.1111/1749-4877.12116

Zischler, H., Höss, M., Handt, O., *et al.* 1995. Detecting dinosaur DNA. *Science*, 268: 1192-1193.

Zuckerkandl, E. i Pauling, L.B. 1962. Molecular disease, evolution, and genic heterogeneity. *In:* Kasha, M. i Pullman, B. (eds). *Horizons in Biochemistry. Academic Press, New York. 189-225.*